DOI: 10.37394/232022.2021.1.1

Pollen Classification of three types of plants of the family *Malvaceae* using Computational Intelligence Approach

VIJAY DHAWALE
Laboratory of Artificial Intelligence, Department of Applied Electronics
Sant Gadge Baba Amravati University
Amravati (Maharashtra)
INDIA

SANJAY DUDUL
Department of Applied Electronics
Sant Gadge Baba Amravati University
Amravati (Maharashtra)
INDIA

JAYKIRAN TIDKE
Laboratory of Reproductive Biology of Angiosperm, Department of Botany
Sant Gadge Baba Amravati University
Amravati (Maharashtra)
INDIA

Abstract: - The *Malvaceae* family is a kind of plants, which are commonly found near habitat places in India. The earlier approaches reported in the literature were tedious and time consuming with less accuracy due to the similarity in shape and the exine sculptures of pollens. We describe a new classification approach for the classification of three types of pollen grains of the *Malvaceae* family based on feature vector comprised of Histogram coefficients and image statistics. The approach presented gives precise accuracy in classification of pollen grains of the same family by using SEM images.

Key-Words: - pollen grain, classifier, SEM, *Malvaceae*, Computational Intelligence, Multi-layer Perceptron Neural Network

1 Introduction

The *Malvaceae* family is a kind of plants that are commonly noticed near habitat places in India. The primary economic uses of malvaceous plants is as a source of natural fibres, the family provides perhaps the world's three most important fibre crops. We have proposed an algorithm to classify pollens of three plants of the *Malvaceae* family based on different performance measures including Average Classification Accuracy (ACA), Mean Squared Error (MSE) and Normalized Mean Squared Error (NMSE).The challenging task in this classification is that there is a great deal of similarity in the shape of all pollens i.e all are circular in shape and other exine features are also almost identical. In view of close similarity among all pollen grains, the result of classification is encouraging.

2 Problem Formulation

In all the three types (taxa) of plants under family of *Malvaceae*, the morphology of pollens is quite similar i.e. similarity in roundness and ornamentation of exine. Hence, we have tested the classifier's performance on such pollens. Thus, the pollen classification problem becomes a three-class classification problem.

We have proposed a novel approach comprising of Image Processing, Feature extraction and a classifier based on Neural Networks for the pollen classification.

3 Importance of Pollen Classification

Pollen studies are widely used in allergy and epidemiology research, fossil fuel exploration, forensic science, food, pharmaceutical, and cosmetic industry, biotechnology and so many other fields. Hence, classification of pollen grains is a challenge that can be effectively accomplished by computational intelligence (CI) approach. The traditional method of pollen classification analyses the pollen morphological characters using microscopy. This procedure was tedious and required experts from the field of palynology.

The work is organized as follows: Section 1 gives brief introduction to the basic concept and literature review for research paper. Section 2 explains the techniques used for development of the CI based pollen classifier. Section 3 summarizes the results produced by classifier and the other parameters. Finally, Section 4 draws the conclusions and also proposes the future scope.

3.1 Literature Survey

Rodriguez-Damian et al. (2004) presented a complete system for classification of pollen allergenic species of Urticaceace family. The images were taken by an optical microscope. A coarse border of pollen grain was estimated using Hough Transform. A set of shape measures was computed, which were used to discriminate between species. They considered 18 images per class, so that total number of images was 234. Similarly, Zhang et al. (2004) used IA texture and shape features to classify to 97% accuracy for 5 taxa of modern New Zealand pollen types. Li et al. (2004) used image analysis (IA) texture features coupled with ANN to correctly classify 100% of 13 taxa of modern pollen and spore types found in New Zealand. Flenley et al. (2004) demonstrated the first successful automated identification, with 100% accuracy. The technique involved a use of neural network classifier applied to surface texture data from LM images. Rodriguez- Damian et al. (2004) proposed brightness and shape descriptors for pollen classification.

Hodgson et al. (2008) proposed the pollen recognition rate of the system, which is accomplished by including grey-level co-occurrence matrix. Carrion et al. (2008) proposed an improved classification of pollen texture images using SVM and MLP. Baladal et al. (2010) suggested a computer vision as its artificial “eye” and an ANN as its artificial “brain”. An automated image analysis procedure was used to extract gray-scale spectral values of pollen image and pollen classifier was designed based on 3-layer ANN and the gray-scale spectral values were used as input. Results showed that the automated procedure correctly classified pollen grains 78.7% of the time. Ticay-Rivas et al. (2011) proposed the combination of features like shape and ornamentation that have been studied earlier and colour features over de-correlated stretched images for enhanced pollen classification by MLP NN based classifier. Nguyen et al. (2013) proposed improved pollen classification with less training efforts by introducing a new selection criterion to obtain the most valuable training samples.

From the literature surveyed, it was found that, most of the approaches based on neural network reported classification accuracy ranging from 90 to 100% and such work was reported by a particular group of researchers and they were working on plant pollen grains found in one particular country i.e. having a geographical boundary. Our approach of CI based classifier supported with neural networks has employed optimal parameters with respect to reduction in time and space complexity. In addition, the approach is supported by new image acquisition techniques such as SEM.

3.2 Design of CI based Classifier

For development of pollen classification algorithm, three different genera (class) are considered and we have collected 12 samples (pollen SEM images) as shown in Table 1. The images of plant species, and its pollen is shown in Appendix. Out of 12 images, 09 images are used for training a classifier and remaining 03 images are used for cross-validation.

Computational intelligence with soft computing approach has been used to develop an algorithm for classification of pollen grain. The SEM images of 03 different pollen species were processed by applying image histogram algorithm. Classifier based on Multilayer Perceptron (MLP) Neural Network is explored for pollen classification. The network is trained using the feature dataset, which is partitioned into training and cross validation (CV) dataset. The neural network is then trained on the training dataset and tested on a different CV dataset to find out the classification accuracy and other performance measures. The performance of classifier was evaluated by comparing MSE, NMSE and ACA on training and CV datasets.

4. Methods

4.1 Pollen Data/Sample Collection

In all the twelve pollen samples of three types of plants from family Malvaceae were available for image dataset. The morphology of pollen is quite similar i.e similar in roundness and ornamentation of exine. Hence, we have tried to test the classifiers’ performance on such pollens. The image-set contains pollen grains of the same type including three different classes (taxa): *Sida acuta* (Bala), *Gossypium hirsutum* (Cotton) and *Thespesia* *populnea* (Paras pipal). Out of the 12 samples available, nine were used for training while three were used for cross-validation dataset.

4.2 Feature/ Image Dataset

The feature extraction technique that was already tested and found efficient in pollen classification was based on Histogram coefficients. Hence, similar feature extraction approach is adopted for the design of classifier.

The dataset used for the classifier includes

1. Histogram coefficients
2. Image Statistics features

The shape descriptor is avoided as it is not seen to provide any useful information due to similarity i.e. circular nature of all pollens, because shape of the pollen is circular for the family *Malvaceae*.

The dataset (features) shown in Table 1 contains 12 instances (exemplars) for three different taxa. The dataset is partitioned into a training and CV dataset. The training dataset constitutes 09 exemplars and remaining 03 exemplars were used for CV.

Table 1. Dataset for pollen grains of family *Malvaceae*

S. No.	Taxa/Class	Shape	Total no. of samples	No. of samples used for training	No. of samples Used for CV
1	*Sida acuta* (Bala)	Circular	05	04	01
2	*Gossypium hirsutum*(Cotton)	Circular	03	02	01
3	*Thespesia populnea* (Paras pipal)	Circular	04	03	01
	Total Samples		12	09	03

4.3 Feature Extraction

For the purpose of classification, the feature dataset is formed with the help of Histogram coefficients along with parameters of image statistics and image texture.

The feature set includes the following features:

i. Statistical features: Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity
ii. Histogram coefficients were extracted by using MATLAB (Mathworks Inc., USA)

Fig.1 and Fig. 2 show typical scatter plots drawn in the input features space and it has been noticed that there exists overlapping of classes among different plant species due to common features shared by different plants of the same family. Nonlinear and complex decision boundaries are required to be estimated for separating out different species.

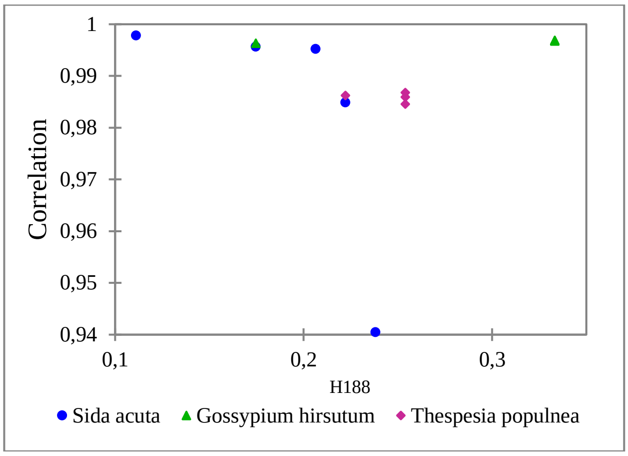

Fig. 1 Scatter plot of feature H188 vs. Correlation

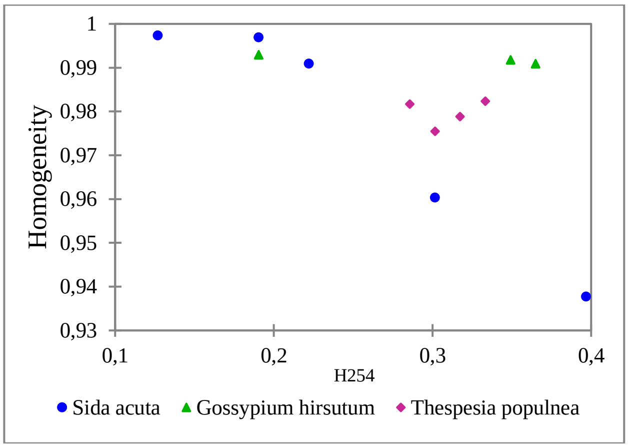

Fig. 2 Scatter plot of feature H254 vs. Homogeneity

4.4 Performance Measures

Performance measures are used to assess the performance of the classifier for a given data set. The performance of the classifier was compared based on MSE, NMSE, Confusion Matrix, Classification Accuracy (CA) and Average Classification Accuracy (ACA) on training and CV datasets.

i. Mean Squared Error (MSE)

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N}(d_{ij}-y_{ij})^2}{NP} \quad \text{-- (1)}$$

Where P = number of output processing elements

N = number of exemplars in the dataset

y_{ij} = network output for exemplar *i* at processing element *j*

d_{ij} = desired output for exemplar *i* at processing element *j*

ii. Normalized MSE (NMSE)

$$NMSE = \frac{PN\,MSE}{\sum_{j=0}^{P} \frac{N \sum_{i=0}^{N} {d_{ij}}^2 - (\sum_{i=0}^{N} d_{ij})^2}{N}} \quad \text{-- (2)}$$

where P = number of output processing elements

N = number of exemplars in the dataset

MSE = mean squared error

d_{ij} = desired output for exemplar *i* at processing element *j*

iii. Confusion Matrix

A confusion matrix is a simple methodology for displaying the classification results of a network. The confusion matrix is defined by labeling the desired classification on the rows and the predicted classifications on the columns.

iv. Classification Accuracy

The classification accuracy indicates the extent to which a CI based classifier is able to accurately classify pollens into three different pollen classes. In other words, the accuracy of the classifier is the degree of closeness of statistical features of pollen to that actual (true) value. In the present study we have taken the classification result in the form of percent accuracy for which maximum or true value is 100.

v. Average Classification Accuracy

ACA is the overall correctness of the result of classifier in terms of percent accuracy for the all three pollen classes and is calculated as the sum of percent correct for a class divided by the total number of classifications.

4.5 Classifier Design

The number of hidden PEs affects the number of connection weights and therefore the training effort per training epoch. The number of training epochs and the number of hidden nodes need to be determined accurately for good classification results. They are determined by parameter optimization. In this case of classification, the feature dataset is formed with the help of Histogram coefficients features along with parameters of image statistics.

The 135 dimensional feature vector, which is to be extracted from the separated *region of interest* (*ROI*) of pollen image, is as follows. F= [H128, H129… H255, Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity]; Where H128, H129… H255 denote the two dimensional discrete Histogram domain coefficients.

Optimal feature vector includes 128 histogram coefficients and seven Image Statistics including Average, Standard Deviation, Entropy, Contrast, Correlation, Energy, Homogeneity.

Following MLP NN based classifier is found to be the best after determining the optimal parameters, selected by systematic experimentations performed and observing the performance measures such as MSE, NMSE and Classification Accuracy.

Variation of PEs in hidden layer#1

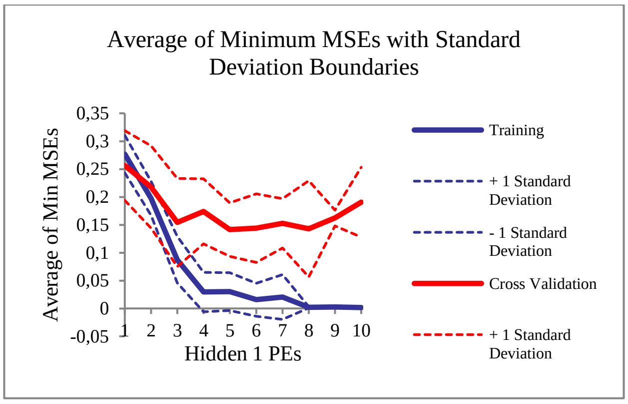

Fig. 3 Determination of number of PEs in Hidden Layer#1 of MLP NN based Classifier

Fig.3 shows variations in average MSE with number of PEs in hidden layer#1. It is noticed that 08 PEs in hidden layer#1 give the minimum MSE. The Characteristics for the MLP NN with single hidden layers are given in Table 2.

Table 2. Characteristic Features of Best MLP NN based Classifier (Hidden Layer#1)

Best Networks	Training	Cross Validation
Hidden 1 PEs	8	8
Run #	1	1
Epoch #	1000	1000
Minimum MSE	5.13477E-05	0.005246745
Final MSE	5.13477E-05	0.005246745

Variation of PEs in hidden layer#2

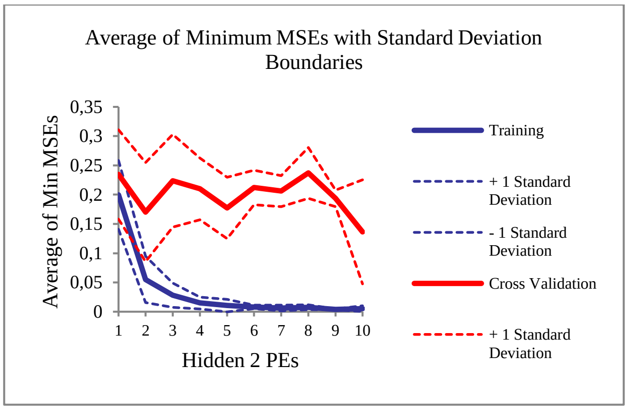

Fig. 4 Determination of number of PEs in Hidden Layer#2 of MLP NN based Classifier

Above figure shows variations in average MSE with number of PEs in hidden layer#2. It is noticed that 10 PEs in hidden layer#2 results into the minimum MSE. The Characteristics for the MLP NN with two hidden layers are given in Table 3.

Table 3. Characteristic Features of Best MLP NN based Classifier (Hidden Layer#2)

Best Networks	Training	Cross Validation
Hidden 2 PEs	10	10
Run #	1	1
Epoch #	1000	1000
Minimum MSE	0.000327677	0.001572327
Final MSE	0.000327677	0.001572327

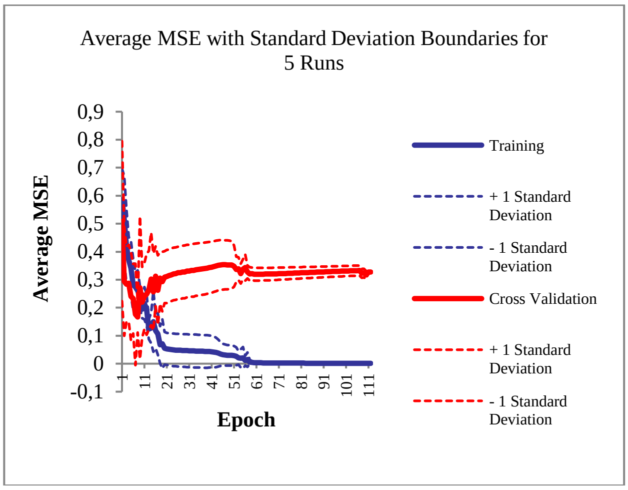

Fig. 5 Variations in Average MSE of MLP NN based Classifier

Above Fig.5 shows the variation of MSE when the classifier is tested on training and CV data. Table 4 and 5 show the performance of the network by observing MSE, NMSE and Classification Accuracy when it is trained on training and CV dataset of three different classes (taxa).

From the experimentation, selected parameters for designing optimum MLP NN (using Histogram coefficients) based classifier are given below: No. of inputs= 12, No. of hidden layers = 02, No. of output PEs = 03, No. of exemplars = 09, Input layer= Tanh/Momentum, No. of Epochs: 120, No. of PEs in hidden layer #1 = 08, No. of PEs in hidden layer #2 = 10. Thus, the architecture of MLP NN based classifier is 12-08-10-3.

4.6 Observation Tables

Table 4 Confusion Matrix and Performance Measures for MLP NN based Classifier (12-08-10-3) on Training Dataset

Output / Desired	**Paras pipal**	**Cotton**	**Bala**
Species(Paras pipal)	3	0	0
Species(Cotton)	0	2	0
Species(Bala)	0	0	4
Performance	Paras pipal	Cotton	Bala
MSE	6.2556E-05	0.000118576	7.04419E-05
NMSE	0.000281502	0.000686044	0.00028529
Percent Correct	100	100	100

Table 5. Confusion Matrix and Performance Measures for MLP NN based Classifier (12-08-10-3) on CV Dataset

Output / Desired	**Paras pipal**	**Cotton**	**Bala**
Species(Paras pipal)	1	0	0
Species(Cotton)	0	1	0
Species(Bala)	0	0	1
Performance	Paras pipal	Cotton	Bala
MSE	5.03454E-05	0.008850914	0.0008899
NMSE	0.000226554	0.039829111	0.00400455
Percent Correct	100	100	100

5 Results

5.1 Analysis

The performance of the pollen classifier is validated meticulously on the basis of important performance measures such as mean squared error (MSE), normalized mean squared error (NMSE), confusion matrix and Average classification accuracy (ACA).

From the results of the classifier performance on the classification of three plant species, it is seen that all the three classes were correctly classified with 100% CA on cross-validation dataset.

The result is very significant because it shows that the pollen classification algorithm is so efficiently designed that even though there is a close similarity in the pollen samples of the same family, it is able to classify the pollen classes on the basis of histogram coefficients and other statistical features.

5.2 Reliability and Robustness (on Uniform and Gaussian Noise)

By observing the performance measures from Table 6, it is found that, the proposed CI based classifier has successfully demonstrated remarkable reliability and robustness in classification of the pollen grains of the family *Malvaceae* as shown in Fig. 6 and 7 . In order to test the reliability and robustness of the classifier against different kinds of noise, such as, Uniform noise and Gaussian noise; controlled amount of both types noise is injected into all inputs of the classifier. For analysis, noise with zero mean value is considered. Subsequently, the variation in the ACA on CV dataset is recorded as variance of the noise is varied from 0 % to 100% in the step of 10%.

It is observed that the proposed classifier has been able to sustain 90% variance of Uniform noise with zero mean as the ACA on CV dataset is maintained at 100 % even after injection of the said noise. It is also seen that the ACA on CV dataset drops from 100 % to 66.66% at 100% variance with zero mean value.

For Gaussian noise, it is confirmed that the proposed classifier can sustain 100 % variance with zero mean as the ACA remains unaltered at 100 % when the variance of injected Gaussian noise is increased from 0 % to 100 %.

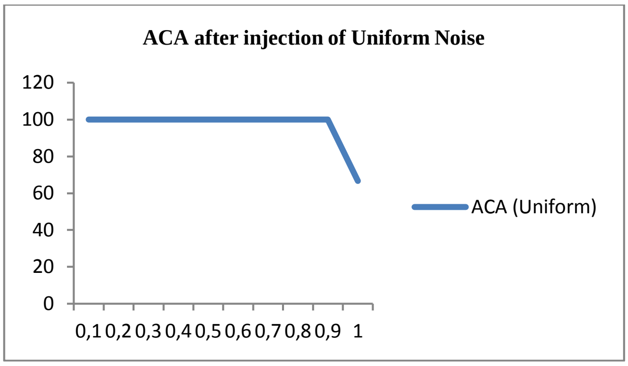

Figure 6. Average Classification Accuracy on CV dataset after injection of Uniform Noise

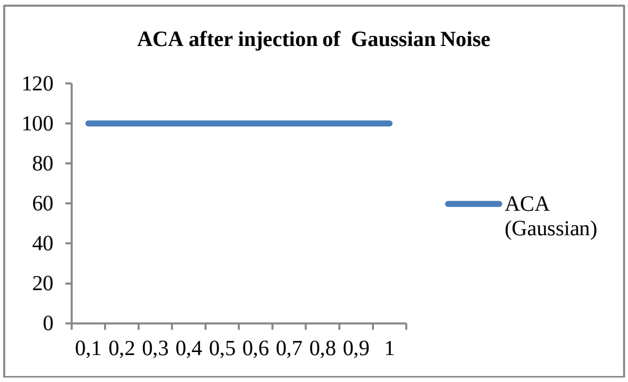

Figure 7. Average Classification Accuracy on CV dataset after injection of Gaussian Noise

6 Conclusion

6.1 Significance of outcomes

A special case study for the pollen species of the same family (*Malvaceae*) has been undertaken, where 12 pollen SEM images were considered. This is a peculiar situation of small-sized dataset. Even in such a situation, our proposed algorithm is found very efficient to classify the pollen taxa with 100% average classification accuracy (ACA) despite the close resemblance of the pollens. Such promising results clearly justify the utility and strength of the proposed algorithm.

The proposed strategy will provide an effective alternative to traditional method of pollen image analysis for plant taxonomy and species identification, which is obvious from the cross-validation performance on the dataset containing three different plant species.

6.2 Future scope

The results obtained are very encouraging and they can form a basis for further evaluation of other available huge volume of pollen grains of the same family of plants from various geographical regions/ places/ locations in the nature.

Acknowledgments

The principal author would like to thank Research laboratories of University Departments of Applied Electronics and Botany at Sant Gadge Baba Amravati University, Amravati along with Vidya Bharati Mahavidyalaya, Amravati for support and all necessary facilities for research purpose. The principal author records special thanks to his research colleague, Mr. Ajay Rajurkar from University Department of Botany for the help in getting pollen SEM images

References:

[1] Erdtman, G. (1952) *Pollen morphology and plant taxonomy. Angiosperms*. Chronica Botanica Co., Waltham, Massachusettes.

[2] Heywood V.H. Scanning Electron Microscopy in the study of plant materials. *Micron*, 1: 1-14,1969

[3] N. R. Nguyen, M.Donalson-Matasci, M.C.Shin, “Improving pollen classification with less training effort”, *IEEE workshop on Applications of Computer Vision (WACV)*, 2013, pp.421 – 426.

[4] E. C. Stillman, J. R. Flenley, "The Needs and Prospects for Automation in Palynology",

Quaternary Science Reviews, vol. 15, Elsevier Science, UK, 1996, pp. 1-5.

[5] Y. Zhang, "Pollen Discrimination Using Image Analysis," Massey University, Palmerston North report, 2001-2003, 2003.

[6] M. Rodriguez-Damian, E. Cernadas, A. Formella, M. Fernandez-Delgado, P. De Sa-Otero, "Automatic detection and classification of grains of pollen based on shape and texture", *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 36, no. 4, pp. 531-542, July 2006 doi: 10.1109/TSMCC.2005.855426

[7] G. P. Allen, R. M. Hodgson, S. R. Marsland, J. R. Flenley,; "Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic objects", *Mechatronics and Machine Vision in Practice*, 2008. M2 VIP 2008. 15th International Conference on, vol., no., pp. 221-226, 2-4 Dec. 2008.

[8] France, I., Duller, A., Duller, G., Lamb, H.: A new approach to automated pollen analysis. Quaternary Science Reviews 18, 537–536, 2000.

[9] Treloar, W.J., Taylor, G.E., Flenley, J.R.: Towards Automation of Palynology 1: Analysis of Pollen Shape and Ornamentation using Simple Geometric Measures, Derived from Scanning Electron Microscope Images. *Journal of Quaternary Science* 19(8), 745–754 (2004).

[10] P. Li and J.R. Flenley. Pollen texture identification using neural networks. *Grana*, pp. 59-64, 1999.

[11] P. Li, W.J. Treloar, J.R. Flenley, L. Empson. Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains, *Journal of Quaternary Science*, vol.19, issue 8, pp.755 – 762, Dec.2004. DOI: 10.1002/jqs.874

[12] K. Holt, G. Allen, R. Hodgson, S. Marsland, J. Flenley. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology*, Vol. 167, Issues 3-4, pp. 175-183, Oct. 2011.

[13] Rodriguez-Damian M., Cernadas E., Formella A., Gonzalez A. “Automatic Identification and Classification of Pollen of The Urticaceae Family”, *Proceedings of Acivs (Advance Concepts for Intelligent Vision Systems)*, Ghent, Belgium, Sept.2-5,2003.

[14] Jones G.D. Pollen Analysis for Pollination Research, Unacetolyzed Pollen, *Journal of Pollination Ecology*, 9(13), 2012, pp 96-107

[15] Jones G.D. and Bryant Jr. V.M., A comparison of pollen counts: Light versus Scanning electron microscopy, *Grana* Vol. 46 (1): 20-33, 2007

[16] Agashe S.N. *Palynology and its applications*. Oxford and IBH Publication Co. Pvt Ltd. New Delhi pp. 16-48, 2006.

[17] Bhattacharya, K., Mujumdar. M.R. and Bhattacharya, S.G. (2006) *A text Book of Palynology*. New Central Book Agency Pvt. Ltd. Kolkata, India.

[18] González, *Digital Image Processing using MATLAB*, 2nd Ed., New Int'l Ed. Ppbk.

[19] Jayaraman S, Veerakumar T, Esakkirajan S, *Digital Image Processing*, McGraw Hill Education, 2009. ISBN:9780070144798

[20] Fausett L., *Fundamentals of Neural Network*, Prentice Hall, 1994

[21] Jain Anil K., *Fundamentals of Digital Image Processing*, PHI Learning Pvt. Ltd., Eastern Economy Edition, New Delhi, 2012.

[22] Haykin Simon, *Neural Networks – A Comprehensive Foundation*, 2nd Edition, McMaster University, Hamilton, Ontario, Canada, Prentice Hall International, Inc., 1994.

Table 6. ACA after injection of Uniform and Gaussian Noise

ACA		ACA (*Malvaceae*) on Histogram coefficients after injection of												
	Noise			Uniform Noise						Gaussian Noise				
CV	Training	Variance	CV	MSE	NMSE	Training	MSE	NMSE	CV	MSE	NMSE	Training	MSE	NMSE
100	100	0.1	100	0.020	0.122	95.13	0.030	0.140	100	0.021	0.098	94.44	0.039	0.195
		0.2	100	0.019	0.086	94.44	0.033	0.164	100	0.012	0.055	91.66	0.057	0.273
		0.3	100	0.033	0.149	94.44	0.043	0.200	100	0.062	0.281	86.11	0.083	0.420
		0.4	100	0.028	0.126	79.16	0.087	0.409	100	0.025	0.114	90.27	0.038	0.171
		0.5	100	0.011	0.049	85.41	0.072	0.346	100	0.023	0.105	87.49	0.073	0.330
		0.6	100	0.005	0.025	86.80	0.059	0.296	100	0.008	0.040	77.08	0.106	0.523
		0.7	100	0.012	0.055	95.83	0.042	0.207	100	0.091	0.410	86.10	0.089	0.458
		0.8	100	0.033	0.150	97.91	0.033	0.163	100	0.015	0.071	89.58	0.052	0.259
		0.9	100	0.031	0.142	79.62	0.090	0.440	100	0.039	0.122	94.44	0.072	0.359
		1	66.66	0.128	0.579	81.24	0.090	0.448	100	0.004	0.020	95.13	0.073	0.169